

University of Groningen

## Improved Diagnostic Validity of the ADOS Revised Algorithms

Oosterling, Iris; Roos, Sascha; de Bildt, Annelies; Rommelse, Nanda; de Jonge, Maretha; Visser, Janne; Lappenschaar, Martijn; Swinkels, Sophie; van der Gaag, Rutger Jan; Buitelaar, Jan

*Published in:*  
Journal of Autism and Developmental Disorders

*DOI:*  
[10.1007/s10803-009-0915-0](https://doi.org/10.1007/s10803-009-0915-0)

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2010

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Oosterling, I., Roos, S., de Bildt, A., Rommelse, N., de Jonge, M., Visser, J., Lappenschaar, M., Swinkels, S., van der Gaag, R. J., & Buitelaar, J. (2010). Improved Diagnostic Validity of the ADOS Revised Algorithms: A Replication Study in an Independent Sample. *Journal of Autism and Developmental Disorders*, 40(6), 689-703. <https://doi.org/10.1007/s10803-009-0915-0>

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

# Improved Diagnostic Validity of the ADOS Revised Algorithms: A Replication Study in an Independent Sample

Iris Oosterling · Sascha Roos · Annelies de Bildt · Nanda Rommelse ·  
Maretha de Jonge · Janne Visser · Martijn Lappenschaar ·  
Sophie Swinkels · Rutger Jan van der Gaag · Jan Buitelaar

Published online: 11 February 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract** Recently, Gotham et al. (2007) proposed revised algorithms for the Autism Diagnostic Observation Schedule (ADOS) with improved diagnostic validity. The aim of the current study was to replicate predictive validity, factor structure, and correlations with age and verbal and nonverbal IQ of the ADOS revised algorithms for Modules 1 and 2 in a large independent Dutch sample ( $N = 532$ ). Results showed that the improvement of diagnostic validity was most apparent for *autism*, except in very young or low functioning children. Results for other autism spectrum disorders were less consistent. Overall, these findings support the use of the more homogeneous revised algorithms, with the use of similar items across developmental cells making it easier to compare ADOS scores within and between individuals.

**Keywords** Autism · ADOS · Algorithm · Sensitivity · Specificity · Diagnosis

## Introduction

The Autism Diagnostic Observation Schedule (ADOS; Lord et al. 2000) is a semi-structured, standardized observational instrument to assess the social and communicative abilities of individuals with possible autism spectrum disorder (ASD). The ADOS consists of four modules intended for use in children, adolescents, and adults with different developmental and language levels. Items are scored from 0 (not abnormal) to 2 or 3 (most abnormal), and a diagnosis of autism or ASD is established if the individual assessed has scores higher than the established cut-off values in the Communication domain, the Social domain, and a sum of the two. The current ADOS diagnostic algorithm does not include items related to repetitive behaviors and restricted interests, although these behaviors are coded if they occur (Lord et al. 2001). Initially, this choice was based on limited available time to notice these kinds of behaviors in the context of the assessment. In research and clinical practice, the ADOS used in combination with the Autism Diagnostic Interview-Revised (ADI-R; Lord et al. 1994) is considered the “gold standard” for diagnosing autism. The ADI-R is a comprehensive, standardized, and semi-structured parent interview that includes items covering all three major domains of dysfunction in autism, namely, quality of reciprocal social interaction, communication, and repetitive, restricted, and stereotyped patterns of behavior.

In 2007, Gotham and colleagues proposed revised algorithms for Modules 1 through 3 to improve the sensitivity and specificity of the ADOS. In the revised algorithms, the original ADOS domains and cut-off values for

I. Oosterling (✉) · S. Roos · N. Rommelse · J. Visser ·  
S. Swinkels · R. J. van der Gaag · J. Buitelaar  
Karakter Child and Adolescent Psychiatry University Center,  
Reinier Postlaan 12, 6525 GC Nijmegen, The Netherlands  
e-mail: i.oosterling@karakter.com

N. Rommelse · M. Lappenschaar · S. Swinkels ·  
R. J. van der Gaag · J. Buitelaar  
Department of Psychiatry, University Medical Center St  
Radboud, Nijmegen, The Netherlands

A. de Bildt  
Child and Adolescent Psychiatry Unit, University Medical  
Centre Groningen, Groningen, The Netherlands

M. de Jonge  
Rudolf Magnus Institute for Neurosciences, Department of Child  
and Adolescent Psychiatry, University Medical Centre Utrecht,  
Utrecht, The Netherlands

N. Rommelse · S. Swinkels · R. J. van der Gaag · J. Buitelaar  
Nijmegen Centre for Evidence-Based Practice (NCEBP),  
Nijmegen, The Netherlands

Social and Communication items have been collapsed into a single factor consisting of 10 items that describes social and communication domain items: the Social Affect factor (SA). In addition, 4 items from a second factor, restricted, totals repetitive behavior (RRB), have been included because RRB domain items may contribute to the diagnosis of autism or ASD, even in the limited context of the ADOS (Lord et al. 2006). There are two diagnostic cut-off scores for the combined SA&RRB domain total, one for autism and one for ASD. In order to reduce ceiling effects in communication items, the revised algorithms distinguish between “Some words” and “No words” in Module 1, on the basis of the item A1 score (overall level of non-echoed language). To reduce the difference between younger more rapidly developing children and older children, the revised algorithms distinguish between age younger and older than 5 years in Module 2.

The new algorithms have important advantages over the original algorithms: they are more homogeneous, which makes comparisons between and within cases across all three modules easier; the effects of age and IQ are probably reduced; and they make direct comparison with ADI-R scores possible. In addition, on the basis of these revised ADOS algorithms, Gotham et al. (2009) recently proposed a calibrated autism severity metric that could prove very useful for identifying trajectories of autism severity in clinical, genetic, and neurobiological research.

To the best of our knowledge, four studies have replicated the revised ADOS algorithms. In the first, small study ( $N = 26$ ), Overton et al. (2008) reported inconsistent results, finding slightly more accurate results for Module 1, although revised algorithm scores for Modules 2 and 3 were similar to those for the original algorithms. In the second study, Gotham et al. (2008) replicated their own study in a large and independent sample ( $N = 1282$ ) and presented sensitivity and specificity data that were similar or better than those for the original algorithms, with the exception of scores for young children with phrase speech and a clinical diagnosis of pervasive developmental disorder not otherwise specified (PDD-NOS). In the third replication study ( $N = 195$ ), Gray et al. (2008) found improved sensitivity with slightly reduced specificity for Module 1/No Words, and improved sensitivity and specificity for Module 1/Some Words. Additional indices of diagnostic accuracy were taken into consideration; however, it was difficult to compare findings with those of the other studies due to dissimilarities in definition of diagnostic groups. In the fourth study, De Bildt et al. (2009) evaluated the revised algorithms in a Dutch, low-functioning sample ( $N = 558$ ). The balance between sensitivity and specificity for Module 2 and 3 was better with the revised algorithms, without there being changes in efficiency of classification. The sensitivity and specificity of

Module 1 showed a more modest improvement, possibly because of the low-functioning sample.

Although these replication studies generally support the use of the revised algorithms, additional research is warranted because the revised algorithms could not be applied to some developmental cells due to limited data (De Bildt et al. 2009; Gotham et al. 2008; Gray et al. 2008). The sample studied by Overton et al. (2008) was too small in general to draw firm conclusions from the results. The sample in the current study did allow for filling up some gaps in the replication literature so far, especially for Module 1/Some Words and both Module 2 cells. In addition, as age and diagnostic groups represented in samples will influence outcome measures at all times, it is of interest to replicate Gotham et al. (2007) findings in an independent sample with a slightly different make-up.

In the current study we sought to augment the findings of earlier studies in a large independent Dutch sample ( $N = 532$ ). The main aim was to examine whether the revised algorithms improved the diagnostic validity of the ADOS. We administered Modules 1 and 2 but not Module 3, because of the limited data available for this module. Additional aims were to study the effects of age and IQ on the revised algorithms, evaluated against the original algorithms, and to verify the factor structure of the revised algorithms.

## Methods

### Participants

Data for 426 participants (aged 15 months to 12 years; 78% male; 95% Dutch Caucasian background) were re-evaluated using the revised algorithms. We included 106 participants with repeated assessments (separated by a mean of  $30.6 \pm 7.0$  months) because Gotham et al. (2007) reported that repeated assessments did not influence findings. Taking the repeated assessments into account, the total sample included 532 cases. Data were provided by three Dutch University Centers for Child and Adolescent Psychiatry, namely, Nijmegen ( $n = 490$ ), Utrecht ( $n = 23$ ), and Groningen ( $n = 19$ ). At least 90% of all participants had been referred for suspected ASD and/or other developmental disorders, most within the context of an extensive early screening project for ASD (see Oosterling et al. 2009). A minority was recruited for research projects (De Bildt et al. 2005; IMGSAC 2001).

On the basis of their use of expressive language, the participants had been assessed with either Module 1 (60%) or Module 2 (40%). They were divided into four developmental cells as suggested in the revised algorithms proposed by Gotham et al. (2007) and as summarized in

Table 1. In order to enhance readability of this paper, the four developmental cells will further be referred to as: cell A, B, C or D (see Table 1).

The clinical diagnoses of the 532 cases were autism (214 cases; 40% of the entire sample), PDD-NOS (134 cases; 25%), Asperger (10 cases; 2%), non-ASD developmental disorders (156 cases; 30%), and no psychiatric disorder (18 cases; 3%). Among the cases with non-ASD developmental disorders, 15 (10%) had non-specific mental retardation, 36 (23%) had language disorders, 52 (33%) had externalizing disorders (Attention Deficit Hyperactivity Disorder or Oppositional Defiant Disorder), 17 (11%) had internalizing disorders (mood or anxiety disorder), and 36 (23%) had other developmental problems. Table 2 provides information on verbal/nonverbal IQ or ratio IQ based on the Psycho-Educational Profile-Revised (PEP-R; Schopler et al. 1990), and on ADOS and ADI-R domain totals.

Compared with the samples of Gotham et al. (2007, 2008) and De Bildt et al. (2009), our sample was *younger* and *higher functioning*, and ADI-R and ADOS mean scores were *lower* or *comparable*. In addition, compared to the other samples, our sample included relatively more non-spectrum and non-autism ASD cases and relatively fewer autism cases. Details about differences in age, (non)verbal IQ, ADI-R, and ADOS mean scores by developmental cell and by diagnostic category are presented in Tables 5 and 6 (Appendices). Our samples appeared to be similar, in terms of age, developmental level, and ADOS and ADI-R mean scores, to the sample administered Module 1 in the study of Gray et al. (2008).

## Measures and Procedure

The diagnostic protocol at the three participating centers included: (a) a psychiatric evaluation, (b) assessment with the ADOS (Lord et al. 2001), (c) a standardized parent interview with the ADI-R (Rutter et al. 2003; available for 70% of the sample), and (d) evaluation of IQ and language skills. In addition, parent–child play was observed for most children aged 0–6 years. A multidisciplinary team comprising minimally a child psychiatrist and a psychologist established the diagnosis, based on DSM-IV-TR (APA

2000) criteria; the child psychiatrist had final responsibility for the diagnosis. Psychologists who met standard requirements for research reliability administered the ADOS and ADI-R. Cognitive abilities were assessed with a variety of instruments, mainly the PEP-R (Schopler et al. 1990; used for 22% of the sample), the Mullen Scales of Early Learning (MSEL; Mullen 1995; used for 31% of the sample), the Snijders-Oomen nonverbal (Dutch) intelligence test (SON-R; Tellegen et al. 1998; used for 29% of the sample), and the Wechsler tests (Wechsler 1997, 2002; used for 9% of the sample). Language abilities were assessed using the Dutch modification of the Reynell Developmental Language Scales (Van Eldik et al. 1995; used for 44% of the sample) and the Schlichting Test for language production (Schlichting et al. 1995; used for 45% of the sample). An indication of verbal IQ was based on either the verbal scale of the MSEL (expressive and receptive language) or the verbal scale of the Wechsler tests or, if these measures were not available, on the mean score on the receptive and expressive scales of the Reynell and Schlichting tests. An indication of nonverbal IQ was based on either the nonverbal scale of the MSEL (visual reception and fine motor scales), or on the SON-R.

## Design and Analyses

For each case, the original algorithms were used to compute Communication and Social domain totals (Lord et al. 2001), and the revised algorithms were used to compute Social Affect (SA) and restricted, repetitive behavior (RRB) domain totals. For both the original and revised algorithms, item scores of 3 were recoded to 2 and item scores of 7 and 8 to 0 as they appear on the algorithm. Diagnostic classifications were then generated using the original or revised threshold cut-offs, both with SA alone and SA&RRB combined (Gotham et al. 2007, p. 621).

To determine whether the revised algorithms improved diagnostic accuracy, we compared sensitivity, specificity, efficiency (correct classification rate), Positive Predictive Value (PPV), and Negative Predictive Value (NPV) of the original and revised algorithms in relation to the clinical classification of each case in the four developmental cells.

**Table 1** Revised algorithm developmental cells

| Age  | Module 1       |                                     | Module 2  |
|------|----------------|-------------------------------------|---|
|      | No speech      | Five or more words - simple phrases | Flexible three words phrases - verbally fluent      |
| <5   | Module 1       | Module 1                            | Module 2  |
|      | No words (A)   | Some words (B)                      | Younger than 5 years (C) <i>n</i> = 108             |
| 5–12 | <i>n</i> = 137 | <i>n</i> = 184                      | Module 2<br>Five years and older (D) <i>n</i> = 103 |

**Table 2** Description of sample characteristics

|                       | Module 1     |      |      |        |                |      |      |        | Module 2                 |       |      |        |                          |      |      |        |
|-----------------------|--------------|------|------|--------|----------------|------|------|--------|--------------------------|-------|------|--------|--------------------------|------|------|--------|
|                       | No words (A) |      |      |        | Some words (B) |      |      |        | Younger than 5 years (C) |       |      |        | Five years and older (D) |      |      |        |
|                       | <i>n</i>     | Mean | SD   | Range  | <i>n</i>       | Mean | SD   | Range  | <i>n</i>                 | Mean  | SD   | Range  | <i>n</i>                 | Mean | SD   | Range  |
| <i>Autism</i>         |              |      |      |        |                |      |      |        |                          |       |      |        |                          |      |      |        |
| Age                   | 104          | 35.6 | 12.9 | 18–73  | 60             | 43.2 | 14.4 | 22–78  | 34                       | 52.3  | 7.6  | 30–60  | 16                       | 79.6 | 26.0 | 60–151 |
| PEP-R Ratio IQ        | 76           | 45.2 | 14.8 | 11–85  | 15             | 59.6 | 8.1  | 47–70  | –                        | –     | –    | –      | –                        | –    | –    | –      |
| VIQ                   | 15           | 43.1 | 20.2 | 13–83  | 33             | 69.7 | 14.9 | 30–102 | 15                       | 85.0  | 17.4 | 44–110 | 13                       | 89.5 | 13.0 | 69–106 |
| NVIQ                  | 19           | 65.6 | 21.7 | 35–107 | 33             | 77.0 | 17.4 | 50–113 | 26                       | 89.9  | 22.0 | 56–131 | 16                       | 87.3 | 20.0 | 50–123 |
| ADI Social            | 88           | 16.7 | 4.7  | 4–26   | 47             | 14.2 | 4.4  | 5–22   | 25                       | 15.9  | 5.2  | 5–25   | 9                        | 13.9 | 6.7  | 5–21   |
| ADI Comm-V            | 1            | 12.0 | –    | –      | 24             | 12.1 | 3.8  | 4–19   | 25                       | 12.2  | 5.3  | 3–23   | 9                        | 10.3 | 3.7  | 6–16   |
| ADI Comm-NV           | 87           | 10.9 | 2.7  | 3–14   | 23             | 8.3  | 2.8  | 4–13   | –                        | –     | –    | –      | –                        | –    | –    | –      |
| ADI RR                | 88           | 3.8  | 1.9  | 0–9    | 47             | 3.7  | 2.3  | 0–8    | 25                       | 5.1   | 3.0  | 0–10   | 9                        | 3.9  | 3.2  | 0–9    |
| ADOS Social           | 104          | 11.8 | 2.6  | 2–14   | 60             | 7.5  | 3.2  | 0–14   | 34                       | 8.0   | 3.6  | 2–16   | 16                       | 5.9  | 3.4  | 0–12   |
| ADOS Comm             | 104          | 5.9  | 1.4  | 2–8    | 60             | 4.0  | 2.1  | 0–9    | 34                       | 4.4   | 2.4  | 0–10   | 16                       | 3.6  | 1.9  | 1–8    |
| ADOS SA               | 104          | 17.0 | 3.2  | 4–20   | 60             | 11.4 | 4.7  | 0–20   | 34                       | 10.5  | 4.8  | 2–18   | 16                       | 8.0  | 4.3  | 0–15   |
| ADOS RRB              | 104          | 3.4  | 1.8  | 0–7    | 60             | 2.6  | 1.9  | 0–7    | 34                       | 1.8   | 1.5  | 0–5    | 16                       | 2.7  | 2.0  | 0–7    |
| <i>Non-Autism ASD</i> |              |      |      |        |                |      |      |        |                          |       |      |        |                          |      |      |        |
| Age                   | 20           | 30.1 | 7.8  | 20–54  | 40             | 36.0 | 6.6  | 26–69  | 41                       | 50.5  | 7.2  | 25–59  | 45                       | 73.2 | 13.3 | 60–113 |
| PEP-R Ratio IQ        | 10           | 56.4 | 12.1 | 36–77  | –              | –    | –    | –      | –                        | –     | –    | –      | –                        | –    | –    | –      |
| VIQ                   | 8            | 73.4 | 19.6 | 50–104 | 36             | 85.6 | 22.1 | 54–178 | 19                       | 98.3  | 15.4 | 70–126 | 35                       | 96.6 | 21.6 | 50–133 |
| NVIQ                  | 9            | 88.6 | 10.3 | 73–103 | 37             | 93.1 | 17.7 | 50–133 | 29                       | 104.2 | 17.3 | 70–148 | 38                       | 91.6 | 22.9 | 50–132 |
| ADI Social            | 15           | 9.6  | 4.6  | 3–18   | 36             | 8.9  | 4.5  | 0–18   | 22                       | 13.6  | 6.7  | 2–29   | 21                       | 9.9  | 5.0  | 3–22   |
| ADI Comm-V            | 2            | 5.5  | 7.8  | 0–11   | 27             | 8.7  | 3.9  | 2–16   | 22                       | 9.6   | 5.2  | 0–18   | 21                       | 6.9  | 3.2  | 2–15   |
| ADI Comm-NV           | 13           | 7.2  | 3.2  | 3–14   | 9              | 6.4  | 2.6  | 2–10   | –                        | –     | –    | –      | –                        | –    | –    | –      |
| ADI RR                | 15           | 2.5  | 2.1  | 0–7    | 36             | 2.7  | 1.9  | 0–8    | 22                       | 4.2   | 3.4  | 0–11   | 21                       | 1.9  | 2.1  | 0–8    |
| ADOS Social           | 20           | 7.0  | 2.8  | 2–13   | 40             | 4.6  | 2.9  | 0–11   | 41                       | 4.1   | 2.8  | 0–11   | 45                       | 5.7  | 2.5  | 1–11   |
| ADOS Comm             | 20           | 4.2  | 1.3  | 2–6    | 40             | 2.5  | 1.6  | 0–7    | 41                       | 2.8   | 1.9  | 0–9    | 45                       | 2.6  | 1.6  | 0–7    |
| ADOS SA               | 20           | 11.0 | 3.7  | 3–18   | 40             | 7.3  | 4.2  | 1–17   | 41                       | 5.8   | 3.7  | 1–15   | 45                       | 7.2  | 3.2  | 1–14   |
| ADOS RRB              | 20           | 2.0  | 1.5  | 0–5    | 40             | 1.1  | 1.1  | 0–4    | 41                       | 1.0   | 1.3  | 0–5    | 45                       | 1.4  | 1.5  | 0–6    |
| <i>Non-ASD</i>        |              |      |      |        |                |      |      |        |                          |       |      |        |                          |      |      |        |
| Age                   | 13           | 26.8 | 7.6  | 15–38  | 84             | 34.5 | 6.7  | 22–66  | 33                       | 50.7  | 6.9  | 35–39  | 42                       | 71.1 | 8.2  | 60–98  |
| PEP-R Ratio IQ        | 4            | 63.5 | 11.2 | 50–77  | 1              | 69.0 | –    | –      | –                        | –     | –    | –      | –                        | –    | –    | –      |
| VIQ                   | 8            | 66.7 | 25.5 | 18–91  | 80             | 87.2 | 18.3 | 38–134 | 21                       | 96.6  | 11.5 | 75–121 | 34                       | 99.5 | 21.6 | 53–145 |
| NVIQ                  | 8            | 89.4 | 14.9 | 56–107 | 81             | 93.1 | 17.3 | 49–140 | 27                       | 98.6  | 15.3 | 66–121 | 38                       | 95.6 | 18.3 | 60–133 |
| ADI Social            | 10           | 7.2  | 4.3  | 1–13   | 67             | 6.8  | 4.5  | 0–18   | 11                       | 6.7   | 4.4  | 1–17   | 19                       | 5.7  | 4.6  | 0–15   |
| ADI Comm-V            | 2            | 2.5  | 2.1  | 1–4    | 49             | 6.2  | 3.3  | 1–13   | 10                       | 4.8   | 3.6  | 1–13   | 19                       | 3.8  | 3.0  | 0–13   |
| ADI Comm-NV           | 8            | 8.3  | 4.0  | 3–14   | 18             | 4.7  | 3.1  | 0–10   | 1                        | 2.0   | –    | –      | –                        | –    | –    | –      |
| ADI RR                | 10           | 1.8  | 1.7  | 0–5    | 67             | 2.3  | 1.8  | 0–8    | 11                       | 1.8   | 1.5  | 0–5    | 19                       | 1.0  | 1.2  | 0–4    |
| ADOS Social           | 13           | 4.8  | 4.1  | 0–13   | 84             | 2.3  | 1.8  | 0–8    | 33                       | 2.3   | 2.1  | 0–8    | 42                       | 2.1  | 1.9  | 0–6    |
| ADOS Comm             | 13           | 2.9  | 1.7  | 0–6    | 84             | 1.4  | 1.2  | 0–5    | 33                       | 1.5   | 1.4  | 0–5    | 42                       | 1.0  | 1.3  | 0–6    |
| ADOS SA               | 13           | 7.4  | 5.5  | 1–18   | 84             | 3.8  | 2.6  | 0–12   | 33                       | 3.3   | 2.7  | 0–11   | 42                       | 3.2  | 2.4  | 0–9    |
| ADOS RRB              | 13           | 1.0  | 1.4  | 0–4    | 84             | 0.6  | 1.0  | 0–4    | 33                       | 0.5   | 1.0  | 0–3    | 42                       | 0.9  | 1.3  | 0–6    |

*Note:* Age = age in months; PEP-R Ratio IQ = (PEP-R developmental age/chronological age) \* 100; VIQ = Verbal IQ; NVIQ = Nonverbal IQ; ADI Social = ADI-R Reciprocal Social Interaction total; ADI Comm-V = ADI-R Communication total for verbal cases; ADI Comm-NV = ADI-R Communication total for nonverbal cases; ADI RR = ADI-R Restricted, Repetitive, and Stereotyped Patterns of Behavior total; ADOS Social = ADOS Reciprocal Social Interaction total (original algorithm); ADOS Comm = ADOS Communication total (original algorithm); ADOS SA = ADOS Social Affect total (revised algorithm); ADOS RRB = ADOS Restricted, Repetitive Behavior total (revised algorithm)

In this analysis, cell A was further divided by mental age as indicated by the PEP-R ( $\leq 15$  months cell A-I and  $\geq 15$  months cell A-II), because Gotham et al. (2007) found that the ADOS is of limited value for children with a mental age of less than 16 months as indicated by a very high percentage of false positive results in this cell (specificity = 0.19). The clinical significance of the indices of diagnostic accuracy were evaluated as:  $< 0.70$  = poor;  $0.70$ – $0.79$  = fair;  $0.80$ – $0.89$  = good;  $0.90$ – $1.00$  = excellent (Cicchetti et al. 1995). We compared the ability of the old and new algorithms to distinguish between: (a) autism and non-spectrum cases, (b) non-autism ASD (PDD-NOS and Asperger's Syndrome) and non-spectrum cases, and (c) ASD and non-spectrum cases. In total 516 valid cases were available for calculation of diagnostic accuracy; 16 cases were excluded because they had missing data for items included in the original or revised algorithms e.g., an '8' on the 'Stereotyped/Idiosyncratic use of words or phrases' item. We also decided to exclude diagnostic categories with fewer than 10 cases per diagnostic category as a small sample size may substantially influence the outcomes. Hence, cell A-I with only 3 non-spectrum cases was removed from this analysis.

In order to accurately interpret ADOS outcomes, we also calculated indices of diagnostic accuracy for the ADI-R, calculated the agreement between ADOS and ADI-R classifications based on Cohen's Kappa (Cicchetti and Sparrow 1981), determined which variables influenced an incorrect ADOS classification, using logistic regression modeling, and determined the distribution of scores for ADOS item A1 (Overall level of Non-echoed Language), to check whether participants received the appropriate module.

To compare the effects of age and (non)verbal IQ on the original and revised algorithm totals, we used T-tests to evaluate the difference between two dependent correlations with  $n - 3^\circ$  of freedom and alpha set at 0.05 (Chen and Popovich 2002). In these analyses the correlations between the new algorithms and age/(non)verbal IQ were compared to the correlation between the old algorithms and age/(non)verbal IQ, while being corrected for the correlation between the new and old algorithms. Cell A was left out of this analysis as verbal IQ and nonverbal IQ could not be established for the majority of the cases in this cell.

To verify the goodness-of-fit of the factor structure of the revised algorithms (SA alone vs. SA&RRB), we applied confirmatory factor analyses (CFA) for categorical data by developmental cell. In these analyses, scores of 8 were marked as missing values and excluded. Items with more than 80% missing values were also omitted. Mplus Version 4.1 software (Muthen and Muthen 1995) was used for these analyses.

## Results

### Improved Diagnostic Validity of the Revised Algorithms?

#### *Autism Versus Non-Spectrum Disorders*

The revised algorithms based on the combined SA&RRB domain total had a clearly better balance between sensitivity and specificity, and between PPV and NPV than the original algorithms, and more cases were correctly classified in cells B, C, and D but not in cell A-II (see Table 3). However, the sensitivity of both the original and revised algorithms was (unacceptably) low in cells B, C, and D (Cicchetti et al. 1995).

#### *Non-Autism ASD Versus Non-Spectrum Disorders*

The diagnostic validity of the revised algorithms based on the SA domain total alone was better than that of the original algorithms in cell A-II and D, whereas it was comparable in cell B and C (see Table 3). In general, the balance between sensitivity and specificity seemed to be better with the revised algorithms (based on SA domain total alone and on the combined SA&RRB domain total); however, in most cells the sensitivity of both the original and revised algorithms was unsatisfactory. In cell C, the PPV of the revised algorithm was considerably lower than that of the original algorithm although it remained above the 0.70 threshold (Cicchetti et al. 1995).

#### *ASD Versus Non-Spectrum Disorders*

The revised algorithms based on SA domain total alone distinguished between ASD (autism and non-autism ASD) cases and non-spectrum cases better (improved sensitivity and efficiency) than the original algorithms. Although the specificity of the SA or combined SA&RRB domain totals of the revised algorithms was lower than that of the original algorithms, it was still higher than 0.70 in all cells. The PPV and NPV were not substantially different between the two algorithms. Only in cell D, the PPV increased a lot with the revised algorithm (SA and SA&RRB combined), and the NPV of the revised algorithm was lower than that of the original algorithm in cell A-II, especially with SA&RRB combined.

#### *Additional Analyses to Explain ADOS Outcomes*

As the diagnostic sensitivity of the ADOS was surprisingly low in all cells regardless of the algorithm used (old or new), we performed additional analyses. First, we calculated the predictive validity of the ADI-R, using cut-



**Table 3** Pearson correlation and *t*-tests for correlations between age(N)VIQ and ADOS totals for the revised and original algorithms

|                                    |  | Meets original algorithm Comm-Soc for ADOS Autism |   |            |  | Meets revised algorithm SARRB for ADOS Autism |             |             |            | Meets revised algorithm SA only for ADOS Autism |       |             |             |            |   |                                      |  |  |                                      |                    |       |        |
|------------------------------------|--|---|---|------------|--|---|-------------|-------------|------------|---|-------|-------------|-------------|------------|---|--------------------------------------|--|--|--------------------------------------|--------------------|-------|--------|
|                                    |  | Sensitivity                                       | Specificity   | Efficiency | PPV  | NPV   | Sensitivity | Specificity | Efficiency | PPV   | NPV   | Sensitivity | Specificity | Efficiency | PPV   | NPV                                  |  |  |                                      |                    |       |        |
| Autism versus Non-spectrum         |  |   |   |            |  |   |             |             |            |   |       |             |             |            |   |                                      |  |  |                                      |                    |       |        |
| Module 1                           | Cell A-II (mental age > 15)<br>(AUT = 52, NS = 10)     | 90  | (97 <sup>1</sup> / 89 <sup>2</sup> )                    | 90         | (91 <sup>1</sup> / 78 <sup>2</sup> )                   | 98  | 64          | 80-96       | 90         | (84 <sup>1</sup> )                              | 100   | 74          | 80-96       | 82         | (89 <sup>1</sup> / 78 <sup>2</sup> )                    | 80                                   | (94 <sup>1</sup> / 80 <sup>2</sup> )                   | 81   | (85 <sup>1</sup> / 71 <sup>2</sup> ) | 95                 | 44    |        |
|                                    | Cell B<br>(AUT = 58, NS = 80)                          | 52  | (88 <sup>1</sup> / 73 <sup>2</sup> / 82 <sup>3</sup> )  | 100        | (96 <sup>1</sup> / 94 <sup>2</sup> / 88 <sup>3</sup> ) | 100   | 74          | 80-96       | 80         | (84 <sup>1</sup> )                              | 100   | 74          | 80-96       | 86         | (86 <sup>1</sup> )                                      | 95                                   | 81   | (93 <sup>1</sup> / 94 <sup>2</sup> / 71 <sup>3</sup> ) | 83                                   | (85 <sup>1</sup> ) | 95    | 78     |
|                                    | Cell C<br>(AUT = 34, NS = 30)                          | 44  | (97 <sup>1</sup> / 85 <sup>2</sup> )                    | 100        | (93 <sup>1</sup> / 100 <sup>2</sup> )                  | 100   | 61          | 72-86       | 70         | (84 <sup>1</sup> )                              | 100   | 61          | 72-86       | 81         | (95 <sup>1</sup> / 81 <sup>2</sup> )                    | 97                                   | (97 <sup>1</sup> / 100 <sup>2</sup> )                  | 78   | (85 <sup>1</sup> )                   | 95                 | 69    |        |
|                                    | Cell D<br>(AUT = 14, NS = 40)                          | 21  | (96 <sup>1</sup> / 63 <sup>2</sup> )                    | 100        | (97 <sup>1</sup> / 92 <sup>2</sup> )                   | 100   | 78          | 88-100      | 88         | (81 <sup>1</sup> )                              | 100   | 78          | 88-100      | 70         | (98 <sup>1</sup> / 84 <sup>2</sup> )                    | 92                                   | 74   | (97 <sup>1</sup> / 100 <sup>2</sup> )                  | 78                                   | (85 <sup>1</sup> ) | 95    | 69     |
|                                    |  | 5-51  |   |            | 66-89  |   | 29-100      | 65-89       |            | 66-89   |       | 29-100      | 65-89       |            | 66-89   |                                      | 29-100   | 65-89  |                                      | 66-89              |       | 29-100 |
| Non-Autism ASD versus Non-spectrum |  |   |   |            |  |   |             |             |            |   |       |             |             |            |   |                                      |  |  |                                      |                    |       |        |
| Module 1                           | Cell A-II (mental age > 15)<br>(N-A ASD = 17, NS = 10) | 88  | (88 <sup>1</sup> )                                      | 60         | (67 <sup>1</sup> )                                     | 79  | 75          | 54-94       | 78         | (71 <sup>1</sup> )                              | 75    | 75          | 54-94       | 74         | (80 <sup>1</sup> )                                      | 80                                   | (76 <sup>1</sup> )                                     | 81   | (88 <sup>1</sup> )                   | 88                 | 73    |        |
|                                    | Cell B<br>(N-A ASD = 40, NS = 80)                      | 43  | (67 <sup>1</sup> / 100 <sup>2</sup> / 86 <sup>3</sup> ) | 89         | (84 <sup>1</sup> / 80 <sup>2</sup> / 63 <sup>3</sup> ) | 65  | 76          | 44-83       | 73         | (71 <sup>1</sup> )                              | 65    | 76          | 44-83       | 68         | (75 <sup>1</sup> / 100 <sup>2</sup> / 86 <sup>3</sup> ) | 81                                   | (79 <sup>1</sup> / 75 <sup>2</sup> / 54 <sup>3</sup> ) | 73   | (66 <sup>1</sup> )                   | 81                 | 80    |        |
|                                    | Cell C<br>(N-A ASD = 41, NS = 30)                      | 37  | (76 <sup>1</sup> / 88 <sup>2</sup> )                    | 97         | (70 <sup>1</sup> / 100 <sup>2</sup> )                  | 94  | 53          | 70-100      | 62         | (61 <sup>1</sup> )                              | 94    | 53          | 70-100      | 59         | (54   | (80 <sup>1</sup> / 88 <sup>2</sup> ) | 73   | (63 <sup>1</sup> / 94 <sup>2</sup> )                   | 62                                   | (73                | 54    |        |
|                                    | Cell D<br>(N-A ASD = 44, NS = 40)                      | 45  | (86 <sup>1</sup> / 56 <sup>2</sup> )                    | 93         | (77 <sup>1</sup> / 64 <sup>2</sup> )                   | 87  | 61          | 66-97       | 68         | (61 <sup>1</sup> )                              | 87    | 61          | 66-97       | 74         | (72 <sup>1</sup> / 62 <sup>2</sup> )                    | 83                                   | (77 <sup>1</sup> / 48 <sup>2</sup> )                   | 77   | (54 <sup>1</sup> )                   | 82                 | 73    |        |
|                                    |  | 30-61   |   | 80-98      | 57-78  |   | 66-97       | 47-73       |            | 57-78   |       | 66-97       | 47-73       |            | 67-93   |                                      | 67-93  |  | 67-93                                |                    | 66-92 | 58-85  |
| ASD versus Non-spectrum            |  |   |   |            |  |   |             |             |            |   |       |             |             |            |   |                                      |  |  |                                      |                    |       |        |
| Module 1                           | Cell A-II (mental age > 15)<br>(ASD = 69, NS = 10)     | 97  |   | 60         |  | 94  | 75          | 84-98       | 92         |   | 94    | 75          | 84-98       | 90         |   | 80                                   |  | 92   |                                      | 97                 | 67    |        |
|                                    | Cell B<br>(ASD = 98, NS = 80)                          | 62  |   | 89         |  | 87  | 66          | 77-94       | 74         |   | 87    | 66          | 77-94       | 72         | (76 <sup>1</sup> )                                      | 86                                   | (96 <sup>1</sup> )                                     | 79   | (83 <sup>1</sup> )                   | 87                 | 72    |        |
|                                    | Cell C<br>(ASD = 75, NS = 30)                          | 52  |   | 97         |  | 96  | 45          | 80-96       | 65         |   | 96    | 45          | 80-96       | 68         |   | 71                                   |  | 71   |                                      | 87                 | 50    |        |
|                                    | Cell D<br>(ASD = 58, NS = 40)                          | 50  |   | 93         |  | 91  | 56          | 80-98       | 67         |   | 91    | 56          | 80-98       | 72         |   | 83                                   |  | 78   |                                      | 86                 | 69    |        |
|                                    |  | 37-63   |   | 57-76      |  | 75-98   | 43-68       |             | 57-76      |   | 75-98 | 43-68       |             | 61-85      |   | 61-85                                |  | 61-85  |                                      | 61-85              |       | 54-81  |

Note: Comm-Soc = ADOS Social-Communication total (original algorithm); SARRB = ADOS Social Affect & Restricted, Repetitive Behaviors (revised algorithm); SA = ADOS Social Affect (revised algorithm); PPV = Positive Predictive Value; NPV = Negative Predictive Value; CI = Confidence Interval. Numbers in parentheses are percentages for different indices as reported by <sup>1</sup> Gotham et al. (2007), <sup>2</sup> Gotham et al. (2008), <sup>3</sup> De Bildt et al. (2009), <sup>4</sup> Gray et al. (2008). Printed in bold = Indices of diagnostic accuracy with poor values (<70; Cicchetti et al. 1995)

off scores for autism and ASD (see Risi et al. 2006, *p.* 1100). In general, compared to the ADOS, the sensitivity of the ADI-R (either for autism vs. non-spectrum or for non-autism ASD vs. non-spectrum) was comparable to lower in cell A and B, and comparable to higher in cell C and D. The specificity of the ADI-R was comparable or slightly lower than that of the ADOS. Second, analysis showed the agreement between the ADOS and the ADI-R classifications to be poor ( $\kappa < 0.40$ ). Third, we ran a binary logistic regression to determine which variables affected incorrect ADOS classification of cases clinically diagnosed with ASD. The independent variables *clinician responsible for final diagnosis* and *psychologist coding the ADOS* did not have a significant influence and were therefore left out from the model. Remaining independent variables that were entered in the model were *age at ADOS administration* and *verbal IQ*. Overall, for the old algorithms, the model reached significance ( $\chi^2(2) = 17.10$ ,  $p < .001$ ), with a significant positive result for ‘verbal IQ’ (Odds Ratio = 1.03, CI = 1.01–1.05,  $p < .001$ ). For the revised algorithms with SA alone and SA&RRB combined the models were also significant (with  $\chi^2(2) = 8.38$ ,  $p < .05$  and  $\chi^2(2) = 25.05$ ,  $p < .001$  respectively), and with a significant positive effect of verbal IQ (Odds Ratio = 1.02, CI = 1.01–1.04,  $p < .01$ , and Odds Ratio = 1.04, CI = 1.02–1.06,  $p < .001$  respectively). Fourth, in order to check that participants had been administered the appropriate module, we looked at the distribution of scores (0, 1, and 2) for ADOS item A1 (Overall Level of Non-echoed Language) in the different developmental cells: it was 0, 0, 58 in cell A-II; 49, 30, 21 in cell B; 95, 5, 0 in cell C; and 95, 4, 1 in cell D. These results indicate that there might have been a ceiling effect in cells C and D (most participants scored 0 on this item), with some cases administered a too easy module.

#### Age and Effects of Verbal and Nonverbal IQ

With the *revised* algorithms (SA alone and SARRB combined) no correlation was found between *age* and ADOS total for cell C. Correlations in all other cells were only low ( $\leq 0.37$ ; see Table 4). Correlations between the *revised* ADOS total (SA alone and SARRB combined) and *(non)verbal or verbal IQ* were low to moderate for all developmental cells (–0.24 to –0.47). In general, age or (non)verbal effects were not *diminished* using the new algorithms with SARRB compared to the old algorithms. With the SA factor alone, however, for some developmental cells reduced effects of age or (non)verbal IQ were found (see Table 4). See Table 7 (Appendix) for more detailed correlations, also with the domain totals of the ADI-R (Lord et al. 1994).

#### Verification of Factor Structure

In confirmatory factor analysis (CFA), a Comparative Fit Index (CFI; Bentler 1990) between 0.9 and 1 indicates good fit. In addition, a root mean square error approximation (RMSEA) of 0.08 or lower is considered satisfactory (Browne and Cudeck 1993). Using these criteria, we found the two-factor model (SA&RRB) of the revised algorithms to fit the data better than the one-factor model (SA) in cells A, B, and C, with CFI values ranging between 0.96 and 1.00 and RMSEAs ranging between 0.04 and 0.08. The goodness-of-fit rating for cell D was slightly poorer (CFI = 0.87, RMSEA = 0.14). Note that in cell A item A3 ‘Intonation’ was scored in only a few cases (11.7%), and was excluded from the analysis.

In their analyses, Gotham et al. (2007) allowed a third factor, called “Joint Attention”, which was consistently confirmed across Modules 1 and 2 in their study. This factor comprised response to joint attention, gesturing, showing, initiating joint attention, and unusual eye contact in cell A, and pointing, gesturing, showing, initiating joint attention, and unusual eye contact in cells B, C, and D. CFA of a three-factor model in the current study was satisfactory for cell A (CFI = 1.00, RMSEA = 0.03), cell B (CFI = 0.96, RMSEA = 0.08), and cell C (CFI = 1.00, RMSEA = 0.05) but less satisfactory for cell D (CFI = 0.89, RMSEA = 0.13). In general, outcomes with the three-factor model were only marginally better than with the two-factor model.

Across modules, 5 items or less had communalities (the percentage of variance in a given item explained by all of the factors) less than 0.4, and only 1 or 2 items had communalities below 0.2, except for cell D that included 4 items with moderate communalities (0.2–0.4) and 4 items with low communalities ( $< 0.2$ ). The communalities for *gestures* and *hand mannerisms* were consistently low or moderate across modules. The number of items with low or moderate communalities did not differ much between the two- or three-factor models. See Tables 8 and 9 (Appendices) for correlations of the 14 items with the two- and three-factor models.

#### Discussion

This study replicates previous findings on the diagnostic validity of the revised algorithms for ADOS Modules 1 and 2, proposed by Gotham et al. (2007), in a relatively large independent Dutch sample. Overall, the children in our sample were younger and higher functioning than those included in previous replication studies and with a different distribution of clinical diagnoses (autism, non-autism ASD, and non-spectrum disorders). For some specific cells, our



**Table 4** Pearson correlation and *t*-tests for correlations between age/(N)VIQ and ADOS totals for the revised and original algorithms

|             | Correlation between revised and original algorithm totals |                      | Revised algorithm |                    | Original algorithm    |                                 |                                    |
|-------------|---|----------------------|-------------------|--------------------|-----------------------|---------------------------------|------------------------------------|
|             | $r_{SA-Comm-Soc}$   | $r_{SARRB-Comm-Soc}$ | $r_{SA-age}$      | $r_{SARRB-age}$    | $r_{Comm-Soc-age}$    | $t_{(SA \text{ vs. } CommSoc)}$ | $t_{(SARRB \text{ vs. } CommSoc)}$ |
| <b>Age</b>  |   |                      |                   |                    |                       |                                 |                                    |
| Cell A-I    | 0.97  | 0.94                 | 0.23              | 0.32*              | 0.33*                 | -3.17**                         | -0.05                              |
| Cell A-II   | 0.98  | 0.94                 | 0.35**            | 0.37**             | 0.37**                | -0.95                           | 0.00                               |
| Cell B      | 0.98  | 0.98                 | 0.29**            | 0.32**             | 0.30**                | -1.15                           | 0.71                               |
| Cell C      | 0.97  | 0.96                 | -0.07             | -0.07              | -0.08                 | 0.24                            | 0.22                               |
| Cell D      | 0.94  | 0.94                 | 0.22*             | 0.31**             | 0.30**                | -2.49*                          | 0.27                               |
|             | $r_{SA-Comm-Soc}$   | $r_{SARRB-Comm-Soc}$ | $r_{SA-(N)VIQ}$   | $r_{SARRB-(N)VIQ}$ | $r_{Comm-Soc-(N)VIQ}$ | $t_{(SA \text{ vs. } CommSoc)}$ | $t_{(SARRB \text{ vs. } CommSoc)}$ |
| <b>VIQ</b>  |   |                      |                   |                    |                       |                                 |                                    |
| Cell B      | 0.98  | 0.98                 | -0.33**           | -0.36**            | -0.36**               | 2.13*                           | 0.24                               |
| Cell C      | 0.97  | 0.96                 | -0.38**           | -0.41**            | -0.44**               | 2.12*                           | 0.92                               |
| Cell D      | 0.94  | 0.94                 | -0.24*            | -0.34**            | -0.29**               | 1.39                            | -1.50                              |
| <b>NVIQ</b> |   |                      |                   |                    |                       |                                 |                                    |
| Cell B      | 0.98  | 0.98                 | -0.36**           | -0.41**            | -0.40**               | 2.33*                           | 0.55                               |
| Cell C      | 0.97  | 0.96                 | -0.32**           | -0.36**            | -0.34**               | 0.77                            | -0.58                              |
| Cell D      | 0.94  | 0.94                 | -0.35**           | -0.47**            | -0.42**               | 2.07*                           | -1.54                              |

Note: SA = Social Affect, SARRB = Social Affect and Restricted, Repetitive Behavior, Comm-Soc = Communication and Reciprocal Social Interaction, (N)VIQ = (non)verbal IQ \*  $p < .05$ , \*\*  $p < .01$

study fills up some gaps in the literature, namely for the Modules 1/Some Words cell and for the Module 2/Younger as well as the Module 2/Older than 5 years cell.

Regarding our main aim, to investigate whether the revised algorithms improved the diagnostic validity of the ADOS compared to the original algorithms, we found: (a) increased predictive validity for *autism cases* with the revised algorithms (combined Social Affect and Restricted, Repetitive Behavior domains—SA&RRB) except for participants who were administered Module 1/No words and had a mental age of > 15 months (cell A-II), (b) less consistent improvement in predictive validity for *non-autism ASD cases*, and (c) increased predictive validity for *ASD cases* as a whole, especially when the revised algorithms based on the SA domain total were used. Previous studies generally indicated that the sensitivity, specificity, and/or efficiency of the revised algorithms were the same or better than those of the original algorithms (De Bildt et al. 2009; Gotham et al. 2007, 2008; Gray et al. 2008; Overton et al. 2008). Our findings support the use of the more homogeneous revised algorithms, with similar items used across developmental cells to allow for easier comparison of ADOS scores within and between individuals.

In previous research, the only exception to the overall finding of improved predictive validity of the revised algorithms was the performance of the ADOS revised

algorithm for non-autism ASD in young children administered Module 2 (cell C). Gotham et al. (2008) reported a marked decrease in sensitivity for non-autism ASD with the revised algorithm based on combined SA&RRB domain total compared with the original algorithm in this cell (0.65 vs. 0.88). In cell C (with 17 non-autism ASD and 18 non-spectrum participants), they reported a high mean score (4) for the ADI-R restricted, repetitive and stereotyped behavior domain but a low ADOS mean score for this domain (RRB score of 1.3), which suggests that the ADOS may miss this type of behavior. We found a similar pattern in cell C, which consisted of 41 individuals with non-autism ASD and 30 individuals with non-spectrum disorders. However, compared to the original algorithms, we did not observe the decrease in sensitivity in our substantially larger sample. This suggests that the revised algorithm also has improved diagnostic validity in children with non-autism ASD who have phrase speech and are younger than 5 years.

The sensitivity of the revised algorithms was much lower in the different developmental cells than that reported in previous studies, whereas the specificity was comparable or even better (De Bildt et al. 2009; Gotham et al. 2007, 2008; Gray et al. 2008). As most cases, including the non-spectrum cases, had been referred for suspected ASD, we would have expected the opposite, namely, a high sensitivity with a lower specificity. The

high specificity of the ADOS found in our study supports the usability of the ADOS for research purposes in order to select narrowly defined groups of participants. However, the low sensitivity is a problem, because it is important not to miss cases. There are a number of possible explanations for this discrepancy in sensitivity with earlier studies:

- (a) In some cells, the sample size per diagnostic group was limited, which may have influenced the results. Gotham et al. (2007, 2008) excluded cells with fewer than 15 participants per diagnostic group, whereas we included some diagnostic groups with only 10 cases (Module 1/No Words, mental age > 15 months) or 14 cases (Module 2, > 5 years) per cell.
- (b) Flaws in ADOS coding could distort results. However, this is improbable because our teams were well trained, and five members are recognized as ADOS trainers by leading institutions in the UK or USA. The trainers supervised less experienced colleagues in order to ensure reliability.
- (c) Sample variation may play a role. Most children in this study were clinically referred within the context of an extensive early screening project (see Oosterling et al. 2009). The fact that our sample was younger, higher functioning, and included fewer core cases of autism than the samples in other studies may have given rise to misclassification. Higher functioning children often show milder ASD symptoms or they may “cover up” their weaknesses in a semi-structured one-to-one context. As the disabilities of higher-functioning ASD children are often more evident in complex situations of daily life, we would expect the ADI-R to be more sensitive, which proved to be the case, although only modestly so, for participants who were administered Module 2.
- (d) Another point concerns how diagnoses were established. As the agreement between ADOS and ADI-R classifications is poor, clinicians must have often relied on only one of the instruments in assigning a DSM-IV diagnosis (APA 2000) in addition to observations from other assessments. As the clinicians who made the final diagnoses were highly experienced and the independent variable ‘clinician responsible for final diagnosis’ was not a predictor of the ADOS false negative cases, there is little reason to question the quality of clinical diagnoses made in this study. However, it is a matter of speculation to what extent differences in ADOS outcomes between the sites included in this study and those in the original study performed in the USA are related to differences in assigning clinical diagnoses (clinician based vs. DSM-IV based). The definition of non-autism ASD might be broader in Europe than in the USA, or the

distinction between autism and non-autism ASD might be slightly different in clinical practice. For this reason, we added the ASD versus non-spectrum disorders comparison in the analyses, which improved diagnostic validity relative to the more strict comparison (autism vs. non-spectrum disorders and non-autism ASD vs. non-spectrum disorders), especially when the revised algorithms based on SA alone were used.

- (e) Finally, the choice of module may have influenced results. Guidelines for module choice are clearly described (Lord et al. 2001) but in clinical practice the choice of whether to administer Module 1 or Module 2 is sometimes difficult and arbitrary. In this regard, Klein-Tasman et al. (2007) found that administration of an ‘easier’ module instead of an appropriate module resulted in under classification of autism in participants administered Module 1/2, and in under classification of ASD in participants administered Module 2/3. Therefore, we sought to verify whether the choice of module could have been responsible for the low sensitivity in our study. Despite careful consideration by the diagnostic team of which module to administer, the distribution of item A1 (Level of Non-echoed Language), with a disproportionately large number of 0 scores, suggests that the choice of module resulted in under classification of at least some cases. This notion is supported by the observation that verbal IQ was higher in incorrectly classified cases than in correctly classified cases.

To recapitulate, while a combination of factors could have led to the relatively low sensitivity in our study, which is a problem, this does not detract from our overall finding that the diagnostic validity of the revised algorithms is better than that of the original algorithms.

Interestingly, the revised algorithms based on a single factor (SA) were better than revised algorithms based on two factors (combined SA&RRB) for identifying cases of non-autism ASD in the youngest/lowest functioning group (cell A-II) and in the oldest/highest functioning group (cell D). Although these findings are not in line with Gotham et al. (2007) and are based on a limited amount of data, at least for cell A-II, they might indicate that restricted and repetitive behavior is of less diagnostic value in very young or low-functioning children (cell A-II) and/or older or higher functioning children at the milder ends of the spectrum (cell D) than in other children.

One of the aims of the revising the ADOS algorithms was to minimize the effect of verbal and nonverbal IQ on ADOS totals (Gotham et al. 2007). However, we found a diminished effect of IQ (verbal and nonverbal) on revised

algorithm totals in only a few cells and only with regard to the SA domain total. This suggests that RRB items are (partly) responsible for the influence of verbal/nonverbal IQ on the combined SA&RRB domain total. This is contrary to previous research (De Bildt et al. 2009; Gotham et al. 2007, 2008), which generally found verbal and nonverbal IQ not to affect the ADOS total calculated with the revised algorithms. The fact that our sample was higher functioning than the samples in previous studies could play a role in this difference.

We found that age did not affect the revised algorithm (SA&RRB) domain total in cell C and had only a minor effect in the other cells; the magnitude of this effect was not smaller with the revised algorithm than with the original algorithm, consistent with the findings of other studies (Gotham et al. 2007, 2008; De Bildt et al. 2009).

We found the two-factor model (SA&RRB) of the revised algorithms to fit the data better than the one-factor model (SA), with a slightly unsatisfactory fit for cell D only. In both independent samples of Gotham et al. (2007, 2008), the two-factor model provided a better fit than the one-factor model in *all* developmental cells. While variation in factor structure is to be expected across samples, the slightly unsatisfactory fit of the model in cell D may be due to the difference in the distribution of diagnoses across samples, which was most apparent in this cell. In our sample, cell D included relatively few autism cases, while autism was overrepresented in cell D in the two studies by Gotham et al. In addition to the two-factor model, Gotham et al. (2007) found a three-factor model, with ‘joint attention’ as the third factor, to fit the data better for both cells A and B, but they decided to use the two-factor model due to its greater consistency across developmental cells. In the current study, the three-factor model was not better in any developmental cell, supporting Gotham et al.’s decision to use the two-factor model.

## Limitations

There are some limitations to this study. We did not include any participants assessed with Module 3, which limited the comparison with previous studies. However, we did fill up gaps in the literature on other modules. Moreover, the small sample size of some cells (A-II and D) limited the interpretation of some results and prevented analysis of algorithm performance in cell A-I. However, in this regard, a new Toddler Module of the ADOS with novel tasks for infants and toddlers has been reported by

Luyster et al. (2009) to meet the diagnostic conditions applicable to very young and/or severely delayed children. A final limitation is that clinical diagnoses were made on the basis of the old algorithms, which may have led to biased outcome measures for the original algorithms. Given this bias, one would have expected the diagnostic agreement to be better with the original algorithms than with the revised algorithms, which was not the case, thereby emphasizing the improved diagnostic validity of the revised algorithms.

## Clinical Implications

In general, the current replication study of ADOS Modules 1 and 2 in a large, independent and well-defined population shows that the revised ADOS algorithms (Gotham et al. 2007) improve diagnostic validity compared with the original algorithms. The improvement in predictive validity was most apparent for *autism*; the sensitivity and specificity of the revised algorithms for *non-autism ASDs* were only marginally better than the original algorithms. The low sensitivity of the revised algorithms may be problematic for some developmental cells or diagnostic subgroups since this reflects increasing discrepancy between the ADOS algorithm and clinical diagnosis. Although the source of the lower sensitivity merits further study, it does once again emphasize that a diagnosis of ASD requires a multidisciplinary approach that includes a variety of assessment measures.

We confirmed the factor structure proposed by Gotham et al. (2007), and the revised algorithms were minimally influenced by age; however, they were not entirely independent of (non)verbal IQ. Overall, our study continues to support the use of the more homogeneous new ADOS algorithms for clinical and research purposes.

**Acknowledgements** We gratefully acknowledge all children and parents who participated in our research project and all clinicians who contributed to the data gathering, especially Nicky de Waal, Tim Woudenberg, Sammy Roording-Ragetlie and Kina Potze. This study was supported by a grant from the Korczak Foundation.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## Appendix

See Tables 5, 6, 7, 8, 9.

**Table 5** Comparison between current and previous samples regarding mean chronological age and (verbal/nonverbal) IQ

|                | Gotham et al. (2007) |    |                |    | Gotham et al. (2008) |    |                |    | De Bildt et al. (2009) |    |                |    |
|----------------|----------------------|----|----------------|----|----------------------|----|----------------|----|------------------------|----|----------------|----|
|                | Autism               |    | Non-autism ASD |    | Autism               |    | Non-autism ASD |    | Autism                 |    | Non-autism ASD |    |
|                | a                    | nv | a              | nv | a                    | nv | a              | nv | a                      | nv | a              | nv |
| Module 1       |                      |    |                |    |                      |    |                |    |                        |    |                |    |
| No words (A)   | <                    | >  | <              | >  | <                    | >  | <              | >  | <                      | >  | <              | >  |
| Some words (B) |                      |    |                |    |                      |    |                |    |                        |    |                |    |
| Module 2       |                      |    |                |    |                      |    |                |    |                        |    |                |    |
| Younger (C)    | <                    | >  | =              | >  | >                    | >  | >              | >  | <                      | >  | <              | >  |
| Older (D)      |                      |    |                |    |                      |    |                |    |                        |    |                |    |

Note: a = chronological age; v = verbal IQ; nv = nonverbal IQ; <= significant difference ( $p < .05$ ) with variable in current sample < variable in previous sample; > = significant difference ( $p < .05$ ) with variable in current sample > variable in previous sample

**Table 6** Comparison between current and previous samples regarding ADI-R and ADOS domain mean scores

| ADI-R domains  | Gotham et al. (2007) |    |                |    | Gotham et al. (2008) |    |                |    | De Bildt et al. (2009) |    |                |    |
|----------------|----------------------|----|----------------|----|----------------------|----|----------------|----|------------------------|----|----------------|----|
|                | Autism               |    | Non-autism ASD |    | Autism               |    | Non-autism ASD |    | Autism                 |    | Non-autism ASD |    |
|                | so                   | cv | cn             | ir | so                   | cv | cn             | ir | so                     | cv | cn             | ir |
| Module 1       |                      |    |                |    |                      |    |                |    |                        |    |                |    |
| No words (A)   | <                    | <  | <              | <  | <                    | <  | <              | <  | <                      | <  | <              | <  |
| Some words (B) |                      |    |                |    |                      |    |                |    |                        |    |                |    |
| Module 2       |                      |    |                |    |                      |    |                |    |                        |    |                |    |
| Younger (C)    | <                    | <  | <              | <  | <                    | <  | <              | <  | <                      | <  | <              | <  |
| Older (D)      |                      |    |                |    |                      |    |                |    |                        |    |                |    |
| ADOS domains   | s                    | co | co             | s  | s                    | co | sa             | rb | s                      | co | sa             | rb |
| Module 1       |                      |    |                |    |                      |    |                |    |                        |    |                |    |
| No words (A)   | <                    | <  | <              | <  | <                    | <  | <              | <  | <                      | <  | <              | <  |
| Some words (B) |                      |    |                |    |                      |    |                |    |                        |    |                |    |
| Module 2       |                      |    |                |    |                      |    |                |    |                        |    |                |    |
| Younger (C)    | <                    | <  | <              | <  | <                    | <  | <              | <  | <                      | <  | <              | <  |
| Older (D)      |                      |    |                |    |                      |    |                |    |                        |    |                |    |

Note: so = ADI-R Reciprocal Social Interaction; cv = ADI-R Communication total for verbal cases; cn = ADI-R Communication total for nonverbal cases; ir = ADI-R Restricted, Repetitive, and Stereotyped Patterns of Behavior; s = ADOS Reciprocal Social Interaction (original algorithm); co = ADOS Communication (original algorithm); sa = ADOS Social Affect (revised algorithm); rb = ADOS Restricted, Repetitive Behavior (revised algorithm)

**Table 7** Correlations between ADOS revised algorithm domain totals and age, (non)verbal IQ, ADI-R domain totals, and ADOS original algorithm domain totals

|                        | Age | VIQ   | NVIQ  | ADI-R  |        | ADOS    |      |        |      |          |      |      |  |
|------------------------|-----|-------|-------|--------|--------|---------|------|--------|------|----------|------|------|--|
|                        |     |       |       | Social | Comm-V | Comm-NV | RR   | Social | Comm | Comm-soc | Play | RR   |  |
| Module 1, No words     |     |       |       |        |        |         |      |        |      |          |      |      |  |
| Mental age ≤ 15 months |     |       |       |        |        |         |      |        |      |          |      |      |  |
| SA total               | r   | –     | –     | 0.46   | –      | 0.36    | 0.29 | 0.97   | 0.68 | 0.97     | 0.67 | 0.55 |  |
|                        | n   | 57    |       | 47     | 0      | 47      | 47   | 57     | 57   | 57       | 57   | 57   |  |
| RRB total              | r   | –     | –     | 0.36   | –      | 0.10    | 0.49 | 0.60   | 0.42 | 0.60     | 0.53 | 0.98 |  |
|                        | n   | 57    |       | 47     | 0      | 47      | 47   | 57     | 57   | 57       | 57   | 57   |  |
| SA & RRB               | r   | –     | –     | 0.48   | –      | 0.30    | 0.41 | 0.94   | 0.66 | 0.94     | 0.70 | 0.78 |  |
|                        | n   | 57    |       | 47     | 0      | 47      | 47   | 57     | 57   | 57       | 57   | 57   |  |
| Module 1 No words      |     |       |       |        |        |         |      |        |      |          |      |      |  |
| Mental age > 15 months |     |       |       |        |        |         |      |        |      |          |      |      |  |
| SA total               | r   | –     | –     | 0.53   | 0.21   | 0.50    | 0.31 | 0.98   | 0.77 | 0.98     | 0.46 | 0.41 |  |
|                        | n   | 80    |       | 66     | 5      | 61      | 66   | 80     | 80   | 80       | 80   | 80   |  |
| RRB total              | r   | –     | –     | 0.22   | 0.64   | 0.01    | 0.34 | 0.38   | 0.37 | 0.40     | 0.33 | 0.97 |  |
|                        | n   | 80    |       | 66     | 5      | 61      | 66   | 80     | 80   | 80       | 80   | 80   |  |
| SA & RRB               | r   | –     | –     | 0.51   | 0.36   | 0.42    | 0.36 | 0.93   | 0.76 | 0.94     | 0.48 | 0.62 |  |
|                        | n   | 80    |       | 66     | 5      | 61      | 66   | 80     | 80   | 80       | 80   | 80   |  |
| Module 1, Some words   |     |       |       |        |        |         |      |        |      |          |      |      |  |
| SA total               | r   | –0.33 | –0.36 | 0.49   | 0.51   | 0.50    | 0.21 | 0.97   | 0.83 | 0.98     | 0.52 | 0.50 |  |
|                        | n   | 184   | 151   | 150    | 100    | 50      | 150  | 184    | 184  | 184      | 184  | 184  |  |
| RRB total              | r   | –0.28 | –0.36 | 0.38   | 0.49   | 0.40    | 0.34 | 0.54   | 0.57 | 0.59     | 0.39 | 0.96 |  |
|                        | n   | 184   | 149   | 150    | 100    | 50      | 150  | 184    | 184  | 184      | 184  | 184  |  |
| SA & RRB               | r   | –0.36 | –0.41 | 0.51   | 0.55   | 0.52    | 0.26 | 0.95   | 0.84 | 0.98     | 0.54 | 0.67 |  |
|                        | n   | 184   | 149   | 150    | 100    | 50      | 150  | 184    | 184  | 184      | 184  | 184  |  |
| Module 2, Younger      |     |       |       |        |        |         |      |        |      |          |      |      |  |
| SA total               | r   | –0.38 | –0.32 | 0.44   | 0.38   | –       | 0.47 | 0.96   | 0.82 | 0.97     | 0.44 | 0.37 |  |
|                        | n   | 108   | 55    | 58     | 57     | 1       | 58   | 108    | 108  | 108      | 108  | 108  |  |
| RRB total              | r   | –0.37 | –0.31 | 0.21   | 0.19   | –       | 0.27 | 0.43   | 0.51 | 0.49     | 0.20 | 0.93 |  |
|                        | n   | 108   | 55    | 58     | 57     | 1       | 58   | 108    | 108  | 108      | 108  | 108  |  |
| SA & RRB               | r   | –0.41 | –0.36 | 0.43   | 0.37   | –       | 0.47 | 0.94   | 0.84 | 0.96     | 0.43 | 0.55 |  |
|                        | n   | 108   | 55    | 58     | 57     | 1       | 58   | 108    | 108  | 108      | 108  | 108  |  |
| Module 2, Older        |     |       |       |        |        |         |      |        |      |          |      |      |  |
| SA total               | r   | –0.24 | –0.35 | 0.38   | 0.46   | –       | 0.29 | 0.92   | 0.79 | 0.94     | 0.23 | 0.36 |  |
|                        | n   | 103   | 82    | 49     | 49     | 0       | 49   | 103    | 103  | 103      | 103  | 103  |  |
| RRB total              | r   | –0.44 | –0.54 | 0.28   | 0.36   | –       | 0.40 | 0.48   | 0.45 | 0.50     | 0.16 | 0.96 |  |
|                        | n   | 103   | 82    | 49     | 49     | 0       | 49   | 103    | 103  | 103      | 103  | 103  |  |
| SA & RRB               | r   | –0.34 | –0.47 | 0.40   | 0.49   | –       | 0.36 | 0.91   | 0.80 | 0.94     | 0.24 | 0.63 |  |
|                        | n   | 103   | 82    | 49     | 49     | 0       | 49   | 108    | 108  | 103      | 103  | 103  |  |

Note: Age = age in months; TIQ = Total IQ; ADI-R Social = ADI-R Reciprocal Social Interaction total; ADI-R Comm-V = ADI-R Communication total for verbal children; ADI-R Comm-NV = ADI-R Communication total for nonverbal children; ADI-R RRB = ADI-R Restricted, Repetitive, and Stereotyped Patterns of Behavior total; ADOS Social = ADOS Reciprocal Social Interaction total (original algorithm); ADOS Comm = ADOS Communication total (original algorithm); ADOS Comm-soc = ADOS Communication and Reciprocal Social Interaction total (original algorithm); ADOS Play = ADOS play total (original algorithm); ADOS RR = ADOS Restricted, Repetitive, and Stereotyped Patterns of Behavior total (original algorithm); SA = ADOS Social Affect total (revised algorithm); RRB = ADOS Restricted, Repetitive Behavior total (revised algorithm). Significant correlations printed in bold ( $p < .05$ ).

**Table 8** Item correlations with the two factor model (SA & RRB)

|                                 |                                   | Module 1, No words ( $N = 137$ ) |                                   | Module 1, Some words ( $N = 184$ ) |                                    | Module 2, Younger ( $N = 108$ ) |                                    | Module 2, Older ( $N = 103$ ) |                                    |
|---------------------------------|-----------------------------------|----------------------------------|-----------------------------------|------------------------------------|------------------------------------|---------------------------------|------------------------------------|-------------------------------|------------------------------------|
| Item                            |                                   | $r$                              | Item                              | $r$                                | Item                               | $r$                             | Item                               | $r$                           | Item                               |
| Social affect                   | B1 Unusual eyecontact             | 0.92                             | B1 Unusual eyecontact             | 0.90                               | B1 Unusual eyecontact              | 0.91                            | B1 Unusual eyecontact              | 0.91                          | B1 Unusual eyecontact              |
|                                 | B4 Gaze and other behaviors       | 0.85                             | B4 Gaze and other behaviors       | 0.93                               | B10 Amount of social communication | 0.91                            | B10 Amount of social communication | 0.91                          | B10 Amount of social communication |
|                                 | B3 Facial expressions             | 0.87                             | B3 Facial expressions             | 0.77                               | B2 Facial expressions              | 0.94                            | B2 Facial expressions              | 0.94                          | B2 Facial expressions              |
|                                 | A2 Frequency of vocalisations     | 0.86                             | A2 Frequency of vocalisations     | 0.87                               | B11 Quality of report              | 0.96                            | B11 Quality of report              | 0.96                          | B11 Quality of report              |
|                                 | B5 Shared enjoyment               | 0.80                             | B5 Shared enjoyment               | 0.78                               | B3 Shared enjoyment                | 0.82                            | B3 Shared enjoyment                | 0.82                          | B3 Shared enjoyment                |
|                                 | B12 Quality of social overtures   | 0.94                             | B12 Quality of social overtures   | 0.92                               | B8 Quality of social overtures     | 0.96                            | B8 Quality of social overtures     | 0.96                          | B8 Quality of social overtures     |
|                                 | B11 Response to joint attention   | 0.76                             | A7 Pointing                       | 0.56                               | A7 Pointing                        | 0.79                            | A7 Pointing                        | 0.79                          | A7 Pointing                        |
|                                 | A8 Gestures                       | 0.58                             | A8 Gestures                       | 0.44                               | A8 Gestures                        | 0.81                            | A8 Gestures                        | 0.81                          | A8 Gestures                        |
|                                 | B9 Showing                        | 0.87                             | B9 Showing                        | 0.55                               | B5 Showing                         | 0.87                            | B5 Showing                         | 0.87                          | B5 Showing                         |
|                                 | B10 Initiation of joint attention | 0.67                             | B10 Initiation of joint attention | 0.59                               | B6 Initiation of joint attention   | 0.79                            | B6 Initiation of joint attention   | 0.79                          | B6 Initiation of joint attention   |
|                                 | A3 Intonation                     | –                                | A5 Stereotyped language           | 0.73                               | A5 Stereotyped language            | 0.82                            | A5 Stereotyped language            | 0.82                          | A5 Stereotyped language            |
|                                 | D1 Unusual sensory interest       | 0.82                             | D1 Unusual sensory interest       | 0.62                               | D1 Unusual sensory interest        | 0.54                            | D1 Unusual sensory interest        | 0.54                          | D1 Unusual sensory interest        |
| Restricted repetitive behaviors | D4 Repetitive interests           | 0.78                             | D4 Repetitive interests           | 0.71                               | D4 Repetitive interests            | 0.86                            | D4 Repetitive interests            | 0.86                          | D4 Repetitive interests            |
|                                 | D2 Hand mannerisms                | 0.30                             | D2 Hand mannerisms                | 0.60                               | D2 Hand mannerisms                 | 0.73                            | D2 Hand mannerisms                 | 0.73                          | D2 Hand mannerisms                 |
| CFI RMSEA                       |                                   | 1.00                             | 0.96                              | 1.00                               | 0.87                               | 0.87                            | 0.87                               | 0.87                          | 0.87                               |
|                                 |                                   | 0.04                             | 0.08                              | 0.05                               | 0.14                               | 0.14                            | 0.14                               | 0.14                          | 0.14                               |

*Note.* Item names have been abbreviated from the Western Psychological Services ADOS item names. Refer to the key from Fig. 6 in the ADOS Manual (Lord et al. 1999) for complete names.  $r$  = correlation; CFI = Comparative Fit Index; RMSEA = Root MeanSquare Error Approximation. For Module 1, No Word item A3 was not applicable due to the high number of missing values (88.3%)



**Table 9** Item correlations with the three factor model (SA, RRB & Joint Attention)

|                                 |                                   | Module 1, No words ( $N = 137$ ) |                                   | Module 1, Some words ( $N = 184$ ) |                                    | Module 2, Younger ( $N = 108$ ) |                                    | Module 2, Older ( $N = 103$ ) |      |
|---------------------------------|-----------------------------------|----------------------------------|-----------------------------------|------------------------------------|------------------------------------|---------------------------------|------------------------------------|-------------------------------|------|
| Item                            |                                   | $r$                              | Item                              | $r$                                | Item                               | $r$                             | Item                               | $r$                           | Item |
| Social affect                   | B1 Unusual eyecontact             | 0.93                             | B1 Unusual eyecontact             | 0.94                               | B1 Unusual eyecontact              | 0.93                            | B1 Unusual eyecontact              | 0.87                          |      |
|                                 | B4 Gaze and other behaviors       | 0.85                             | B4 Gaze and other behaviors       | 0.93                               | B10 Amount of social communication | 0.91                            | B10 Amount of social communication | 0.72                          |      |
|                                 | B3 Facial expressions             | 0.88                             | B3 Facial expressions             | 0.77                               | B2 Facial expressions              | 0.94                            | B2 Facial expressions              | 0.64                          |      |
|                                 | A2 Frequency of vocalisations     | 0.87                             | A2 Frequency of vocalisations     | 0.87                               | B11 Quality of report              | 0.96                            | B11 Quality of report              | 0.96                          |      |
|                                 | B5 Shared enjoyment               | 0.81                             | B5 Shared enjoyment               | 0.78                               | B3 Shared enjoyment                | 0.82                            | B3 Shared enjoyment                | 0.63                          |      |
|                                 | B12 Quality of social overtures   | 0.95                             | B12 Quality of social overtures   | 0.92                               | B8 Quality of social overtures     | 0.97                            | B8 Quality of social overtures     | 0.96                          |      |
|                                 | B11 Response to joint attention   | 0.76                             | A7 Pointing                       | 0.58                               | A7 Pointing                        | 0.80                            | A7 Pointing                        | 0.67                          |      |
|                                 | A8 Gestures                       | 0.58                             | A8 Gestures                       | 0.46                               | A8 Gestures                        | 0.84                            | A8 Gestures                        | 0.51                          |      |
|                                 | B9 Showing                        | 0.87                             | B9 Showing                        | 0.58                               | B5 Showing                         | 0.90                            | B5 Showing                         | 0.52                          |      |
|                                 | B10 Initiation of joint attention | 0.67                             | B10 Initiation of joint attention | 0.62                               | B6 Initiation of joint attention   | 0.80                            | B6 Initiation of joint attention   | 0.47                          |      |
|                                 | A3 Intonation                     | –                                | A5 Stereotyped language           | 0.73                               | A5 Stereotyped language            | 0.82                            | A5 Stereotyped language            | 0.62                          |      |
|                                 | D1 Unusual sensory interest       | 0.82                             | D1 Unusual sensory interest       | 0.62                               | D1 Unusual sensory interest        | 0.54                            | D1 Unusual sensory interest        | 0.45                          |      |
| Restricted repetitive behaviors | D4 Repetitive interests           | 0.78                             | D4 Repetitive interests           | 0.72                               | D4 Repetitive interests            | 0.86                            | D4 Repetitive interests            | 0.94                          |      |
|                                 | D2 Hand mannerisms                | 0.30                             | D2 Hand mannerisms                | 0.60                               | D2 Hand mannerisms                 | 0.73                            | D2 Hand mannerisms                 | 0.67                          |      |
| CFI RMSEA                       |                                   | 1.00                             | 0.96                              | 1.00                               | 0.89                               |                                 |                                    |                               |      |
|                                 |                                   | 0.03                             | 0.08                              | 0.05                               | 0.13                               |                                 |                                    |                               |      |

*Note.* Item names have been abbreviated from the Western Psychological Services ADOS item names. Refer to the key from Fig. 6 in the ADOS Manual (Lord et al. 1999) for complete names.  $r$  = correlation; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error Approximation. For Module 1, No Word item A3 was not applicable due to the high number of missing values (88.3%)

## References

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders, text revision* (Fourth Edition ed.). Washington, DC: American Psychiatric Association.
- Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, 107, 238–246.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.
- Chen, P. Y., & Popovich, P. M. (2002). *Correlation: Parametric and nonparametric measures*. Thousand Oaks, CA: Sage Publications.
- Cicchetti, D. V., & Sparrow, S. S. (1981). Developing criteria for establishing inter-rater reliability of specific items in a given inventory. *American Journal of Mental Deficiency*, 86, 127–137.
- Cicchetti, D. V., Volkmar, F., Klin, A., & Showalter, D. (1995). Diagnosing autism using ICD-10 criteria: A comparison of neural networks and standard multivariate procedures. *Child Neuropsychology*, 1, 26–37.
- De Bildt, A., Sytema, S., Kraijer, D., & Minderaa, R. (2005). Prevalence of pervasive developmental disorders in children and adolescents with mental retardation. *Journal of Child Psychology and Psychiatry*, 46, 275–286.
- De Bildt, A., Sytema, S., Van Lang, N. D. J., Minderaa, R. B., Van Engeland, H., & De Jonge, M. V. (2009). Evaluation of the ADOS revised algorithm: The applicability in 558 Dutch children and adolescents. *Journal of Autism and Developmental Disorders*, 39(9), 1350–1358.
- Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 39, 693–705.
- Gotham, K., Risi, S., Dawson, G., Tager-Flusberg, H., Joseph, R., Carter, A., et al. (2008). A replication of the Autism Diagnostic Observation Schedule (ADOS) revised algorithms. *American Academy of Child and Adolescent Psychiatry*, 47(6), 642–651.
- Gotham, K., Risi, S., Pickles, A., & Lord, C. (2007). The autism diagnostic observation schedule: revised algorithms for improved diagnostic validity. *Journal of Autism and Developmental Disorders*, 37, 613–627.
- Gray, K. M., Tonge, B. J., & Sweeney, D. J. (2008). Using the Autism diagnostic interview- revised and the autism diagnostic observation schedule with young children with developmental delay: Evaluating diagnostic validity. *Journal of Autism and Developmental Disorders*, 38, 657–667.
- International Molecular Genetic Study of Autism Consortium (IMG-SAC). (2001). A genomewide screen for autism: strong evidence for linkage to chromosomes 2q, 7q, and 16p. *American Journal of Human Genetics*, 69(3), 570–581.
- Klein-Tasman, B. P., Risi, S., & Lord, C. E. (2007). Effect of language and task demands on the diagnostic effectiveness of the Autism Diagnostic Observation Schedule: the impact of module choice. *Journal of Autism and Developmental Disorders*, 37, 1224–1234.
- Lord, C., Risi, S., DiLavore, P., Shulman, C., Thurm, A., & Pickles, A. (2006). Autism from two to nine. *Archives of General Psychiatry*, 63(6), 694–701.
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., et al. (2000). The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30(3), 205–223.
- Lord, C., Rutter, M., DiLavore, P., & Risi, S. (1999). *Autism diagnostic observation schedule (ADOS), manual*. Los Angeles, CA: Western Psychological Services.
- Lord, C., Rutter, M., DiLavore, P., & Risi, S. (2001). *Autism diagnostic observation schedule (ADOS) manual*. Los Angeles, CA: Western Psychological Services.
- Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview-Revised: A revised version of the diagnostic Interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24, 659–685.
- Luyster, R., Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K., Bishop, S., Esler, A., Hus, V., Oti, R., Richler, J., Risi, S., & Lord, C. (2009). The Autism Diagnostic Observation Schedule-Toddler Module: A new module of a standardized diagnostic measure for autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 39(9), 1305–1320.
- Mullen, E. (1995). *Mullen scales of early learning* (AGS Edition ed.). Circle Pines, MN: American Guidance Service.
- Muthen, L. K., & Muthen, B. O. (1995). *M-plus user's guide*. Los Angeles, CA: Muthen and Muthen.
- Oosterling, I. J., Wensing, M., Swinkels, S. H., van der Gaag, R. J., Visser, J. C., Woudenberg, T., Minderaa, R., Steenhuis, M. P., & Buitelaar, J. K. (2009). Advancing early detection of autism spectrum disorder by applying an integrated two-stage screening approach. *Journal of Child Psychology and Psychiatry*. doi: 10.1111/j.1469-7610.2009.02150.x.
- Overton, T., Fielding, C., & Garcia de Alba, R. (2008). Brief report: Exploratory analysis of the ADOS revised algorithm: Specificity and predictive value with Hispanic children referred for autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 38, 1166–1169.
- Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., et al. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, 45, 1094–1103.
- Rutter, M., Le Couteur, A., & Lord, C. (2003). *ADI-R: Autism diagnostic interview-Revised*. Los Angeles, CA: Western Psychological Services.
- Schlichting, J. E. P. T., van Eldik, M. C. M., Lutje Spelberg, H. C., Van der Meulen, S. J., & Van der Meulen, B. F. (1995). *Schlichting test voor taalproductie*. Lisse: Swets Test Publishers.
- Schopler, E., Reichler, R. J., Bashford, A., Lansing, M. D., & Marcus, L. M. (1990). *The psychoeducational profile revised (PEP-R)*. Austin: Pro-Ed.
- Tellegen, P. J., Winkel, M., Wijnberg-Williams, B. J., & Laros, J. A. (1998). *Snijders-Oomen Niet-verbale Intelligentietest SON-R 2.5–7*. Lisse: Swets & Zeitlinger.
- van Eldik, M. C. M., Schlichting, J. E. P. T., Lutje Spelberg, H. C., van der Meulen, B. F., & van der Meulen, S. J. (1995). *Reynell test voor taalbegrip*. Amsterdam: Harcourt Test Publishers.
- Wechsler, D. (1997). *Wechsler preschool and primary scale of intelligence-revised*. Lisse: Swets & Zeitlinger.
- Wechsler, D. (2002). *Wechsler intelligence scale for children-III.-NL*. Amsterdam: Harcourt Test Publishers.